OCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

(12)

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 1992	memorandum

4. TITLE AND SUBTITLE
Recognition and Structure from one 2D Model View: Observations on Prototypes, Object Classes and Symmetries

5. FUNDING NUMBERS
IRI-8719394
N00014-89-J-3139
8814612-MIP
N00014-91-J-4038

6. AUTHOR(S)
Tomaso POggio and Thomas Vetter

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Artificial Intelligence Laboratory
545 Technology Square
Cambridge, Massachusetts 02139

8. PERFORMING ORGANIZATION REPORT NUMBER
AIM 1347
C.B.I.P. 69

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
Information Systems
Arlington, Virginia 22217

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
None

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution of this document is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

According to the *1.5 views theorem* (Poggio, 1990; Ullman and Basri, 1991) recognition of a specific 3D object (defined in terms of pointwise features) from a novel 2D view can be achieved from at least two 2D model views (in the data basis, for each object, for orthographic projection). In this note we discuss how recognition can be achieved from a single 2D model view. The basic idea is to exploit transformations that are specific for the object class corresponding to the object – and that may be known a priori or may be learned from views of other "prototypical" objects of the same class – to generate new model views from the only one available. The paper is organized in two distinct parts. In the first part, we discuss how to exploit prior knowledge of an object's symmetry. We prove that for any bilaterally symmetric 3D object one non-accidental 2D model view is sufficient for recognition. We also prove

(continued on back)

14. SUBJECT TERMS (key words)
object recognition, symmetry
class and prototypes, learning

15. NUMBER OF PAGES
21

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED

Block 13 continued:

that for bilaterally symmetric objects the correspondence of four points between two views determines the correspondence of *all* other points. Symmetries of higher order allow the recovery of structure from one 2D view. In the second part of the paper, we study a very simple type of object classes that we call *linear object classes*. Linear transformations can be learned exactly from a small set of examples in the case of linear object classes and used to produce new views of an object from a single view. We also provide natural examples of linear object classes induced by symmetry properties of the objects.

(12)

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
and
CENTER FOR BIOLOGICAL INFORMATION PROCESSING
WHITAKER COLLEGE

A.I. Memo No. 1347 February 1992
C.B.I.P. Paper No. 69

Recognition and Structure from one 2D Model View: Observations on Prototypes, Object Classes and Symmetries

Tomaso Poggio and Thomas Vetter

Abstract

According to the *1.5 views theorem* (Poggio, 1990; Ullman and Basri, 1991) recognition of a specific 3D object (defined in terms of pointwise features) from a novel 2D view can be achieved from at least two 2D model views (in the data basis, for each object, for orthographic projection). In this note we discuss how recognition can be achieved from a single 2D model view. The basic idea is to exploit transformations that are specific for the object class corresponding to the object – and that may be known a priori or may be learned from views of other "prototypical" objects of the same class – to generate new model views from the only one available. The paper is organized in two distinct parts. In the first part, we discuss how to exploit prior knowledge of an object's symmetry. We prove that for any bilaterally symmetric 3D object one non-accidental 2D model view is sufficient for recognition. We also prove that for bilaterally symmetric objects the correspondence of four points between two views determines the correspondence of *all* other points. Symmetries of higher order allow the recovery of structure from one 2D view. In the second part of the paper, we study a very simple type of object classes that we call *linear object classes*. Linear transformations can be learned exactly from a small set of examples in the case of linear object classes and used to produce new views of an object from a single view. We also provide natural examples of linear object classes induced by symmetry properties of the objects.

This paper describes research done within the Center for Biological Information Processing in the Department of Brain and Cognitive Sciences, and at the Artificial Intelligence Laboratory. This research is sponsored by grants from the Office of Naval Research, Cognitive and Neural Sciences Division; by a grant from the National Science Foundation under contract IRI-8719394; by the Artificial Intelligence Center of Hughes Aircraft Corporation (LJ90-074); by Office of Naval Research contract N00014–89–J–3139 under the DARPA Artificial Neural Network Technology Program; and by NSF and DARPA under contract 8814612-MIP. Support for the A.I. Laboratory's artificial intelligence research is provided by ONR contract N00014–91–J–4038. Tomaso Poggio is supported by

the Uncas and Helen Whitaker Chair at the Whitaker College, Massachusetts Institute of Technology. Thomas Vetter holds a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft (Ve 135/1-1).

1 Introduction

Techniques have been recently developed that can learn to recognize a specific 3D object after a "learning" stage in which a few 2D views of the object are used as training examples (Poggio and Edelman, 1990; Edelman and Poggio, 1990). A lower bound on the number of views is provided by the *1.5 view theorem* (Poggio, 1990; see also Ullman and Basri, 1991 who pioneered the linear combination approach and Huang and Lee, 1989) that implies that 2 views - appropriately defined - may be sufficient in the orthographic case. Under more general conditions (perspective projection, more general definition of view, non uniform transformations etc.) many more views may be required (Poggio and Edelman's estimate is in the order of 100 for the whole viewing sphere using their approximation network).

Though this is an easily satisfied requirement in many cases, there are situations in which only one 2D view is available as a model. As an example, consider the problem of recognizing a face from just one view: humans can do it, even for different facial expressions (of course an almost frontal view may not be sufficient for recognizing a profile view and in fact the praxis of person identification requires usually a frontal *and* a side view).

Clearly one single view of a generic 3D object (if shading is neglected) does not contain sufficient 3D information. If, however, the object belongs to a class of similar objects (prototypes), it seems possible to infer appropriate transformations for the class and use them to generate other views of the specific object from just one 2D view of it. We are certainly able to recognize faces which are slightly rotated from just one quasi-frontal view, presumably because we exploit our extensive knowledge of the typical 3D structure of faces.

One can pose the following problem: *is it possible from one 2D view of a 3D object to generate other views, exploiting knowledge of the legal transformations associated with objects of the same class?* A positive answer would imply (for orthographic projection and uniform affine transformations) that a novel 2D view may be recognized from a single 2D model view, because of the 1.5 views theorem [1].

This note is divided in two distinct parts. In the *first* part we consider the case in which legal transformations for a specific object (i.e. transformations that generate new correct views from a given one) are immediately available as a property of the class. In particular, we will discuss certain symmetry properties. In the *second* part, we consider the problem of learning appropriate transformations from examples of other objects of the same class.
The main results in the first part of the paper are:

1. we prove that for any bilaterally symmetric 3D object (such as a face) one 2D model view is sufficient for recognition of a novel 2D view (for orthographic projection and uniform affine transformations). This result is equivalent to the following statement: for bilaterally symmetric objects a model based recognition invariant (as defined by Weinshall, 1992) can be learned from just one model 2D view;

2. we also prove that for symmetries of higher order (such as two-fold symmetries, i.e. bilateral symmetry with respect to two symmetry planes) it is possible to recover structure from one 2D view.

[1] A positive answer would also make possible the use of other recognition techniques such as Poggio and Edelman's technique - and its extensions, possibly including the correlation based version (Brunelli and Poggio, 1991) - by using the newly generated views as a training set.

In the second part of the paper we first argue that transformations that generate additional model views from a single view may be learned at least approximatively from examples of objects of the same class. We then

1. introduce the definition of "linear classes",
2. show that for linear classes one 2D model view is sufficient to generate exact additional views (and therefore to perform recognition of a novel view);
3. discuss examples of linear classes and prove that object symmetries induce a natural set of linear classes: for instance, bilaterally symmetric objects are a linear class.

In the final section,, we briefly mention some the implications of our results for the practical recognition of bilaterally symmetric objects such as faces, for human perception of 3D structure from single views of geometric objects and, more generally, for the role of symmetry detection in human vision.

2 PART I: Object Symmetries Recover Recognition and Structure from One 2D View

2.1 Recognition from One 2D Model View

Suppose that we have a model 2D view of an object. Assume further that (a) we know *a priori* that the object is bilaterally symmetric (for instance because we identify the class to which it belong and we know that this class has the property of bilateral symmetry) and (b) we know a pair of symmetric points in the 2D view. For the purpose of this first part we define an object to be *bilaterally symmetric* if the following transformation of any 2D view of a pair of symmetric points of the object yields a *legal view* of the pair, that is the orthographic projection of a rigid rotation of the object

$$D\mathbf{x}_{pair} = \mathbf{x}^*_{pair} \tag{1}$$

with

$$\mathbf{x}_{pair} = \begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{pmatrix} \qquad \mathbf{x}^*_{pair} = \begin{pmatrix} -x_2 \\ -x_1 \\ y_2 \\ y_1 \end{pmatrix}$$

and

$$D = \begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Notice that symmetric pairs are the elementary features in this situations and points lying on the symmetry plane are degenerate cases of symmetric pairs. Notice also that our definition

of symmetry is in the same spirit as its use in physics, where symmetries of an abstract object are typically defined in terms of properties of the object under an appropriate set of transformations.

Geometrically, this simply means that for bilaterally symmetric objects simple transformations of a 2D view yield other views that are *legal*. The transformations are similar to mirroring one view around an axis in the image plane, as shown in Figure 1 top (where the left image is "mirrored" into the right one) and correspond – but only for a bilaterally symmetric object – to proper rotations of a rigid 3D object and their ortographic projection on the image plane.

Equation 1 defines one such transformation and generates an additional view from the one model view (and the knowledge of bilateral symmetry). The 2 views $\mathbf{x}_{pair}$ and $\mathbf{x}^*_{pair}$ are linearly independent, unless $\mathbf{x}_{pair} = \lambda \mathbf{x}^*_{pair}$, which is equivalent to the condition that $\mathbf{x}_{pair}$ is the solution of the eigenvalue problem

$$D\mathbf{x}_{pair} = \lambda \mathbf{x}_{pair},$$

that is, unless $\mathbf{x}_{pair}$ is a view which is left invariant(modulus a sign) by the symmetry operation D. The eigenvalue problem has exactly two solutions (with $\lambda = \pm 1$) which correspond to "accidental" views such as a perfectly frontal view, an exact side view, and reflection about the image plane (obtainable for "transparent" objects by a π rotation). These $\mathbf{x}$ are the only ones for which the symmetry operation D does not provide a linearly independent new view. The same argument can be repeated for all symmetric pairs (points on the symmetry axes are of course a degenerate case of a pair) and all transformations.

Thus, bilateral symmetry allows the generation of an additional, linearly independent view of the object. The 1.5 views theorem (see Appendix A.5) can then be used to compute the 3D basis that spans the spaces V_x and V_y of the object. Recognition of any view of the object is then possible. We have thus proved

Theorem 2.1 *A single 2D view of a bilateral symmetric object (containing at least 2 symmetric, nondegenerate pairs, once translations are factored out) yields a three dimensional basis for the vector spaces V_x and V_y provided that the view is not an "accidental" view, i.e. is not a solution of $D\mathbf{x} = \pm\mathbf{x}$.*

Notice that bilateral symmetry provides from one 2D view a total of eight 2D views, each corresponding to a different rotation of the original 3D view. Four of the eight views are linearly independent (two linearly independent vectors of the x coordinates and two for the y coordinates). [2] Moses and Ullman (1991) derived a result about recognition functions of symmetric objects that is consistent with our theorem and complements it.

Notice that it is also possible to define bilateral symmetry for the 3D object and then show that this definition yields the one above in the following way. Let us call a *3D object bilaterally symmetric* if there exist a position and orientation of the object relative to a given 3D cartesian coordinate system for which each feature point

[2] The depth ambiguity of any 2D view of "transparent" objects correspondes to a rotation of a single rigid object (notice in the non symmetric case the two views cannot be interpreted as a rotation of a rigid object)

$$\mathbf{x}_1 = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix},$$

has either $x = 0$ or a symmetric point $\mathbf{x}_2$, such that

$$\mathbf{x}_2 = \begin{pmatrix} -x_1 \\ y_1 \\ z_1 \end{pmatrix}.$$

It is easy to verify that there is a rotation ψ around the z axis in 3D under which the vector of the coordinates of the two symmetric points transforms into

$$\mathbf{x}_{pair} = \begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{pmatrix},$$

whereas a rotation of $-\psi$ maps it into

$$\mathbf{x}^*_{pair} = \begin{pmatrix} -x_2 \\ -x_1 \\ y_2 \\ y_1 \end{pmatrix}$$

.

2.1.1 A Recognition Algorithm

A single 2D model view together with knowledge that the object is bilaterallly symmetric can be used for recognition (in the same spirit as Ullman and Basri, 1991) in the following way.

1. Take $\mathbf{x}_1$ and $\mathbf{y}_1$ (the vectors of the x and y coordinates of the n feature points) from the available view and generate a third vector $\mathbf{x}_2$ (or $\mathbf{y}_2$) by applying the symmetry transformation D to $\mathbf{x}_1$ (or $\mathbf{y}_1$).

2. Make a $2n \times 6$ matrix B with its 6 columns representing a basis for $V^{2N}_{obj} = V^N_x \oplus V^N_y$. An explicit form of B is

$$B = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{y}_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{y}_1 \end{pmatrix}$$

3. Check that B is full rank (for instance $(B^T B)^{-1}$ exists). If B is not full rank try others of the legal views induced by symmetry.

4. A novel view $\mathbf{t}$ (we assume here that the first n components are the x coordinates followed by n y) of the same object must be in the space spanned by the columns of B, and therefore must satisfy

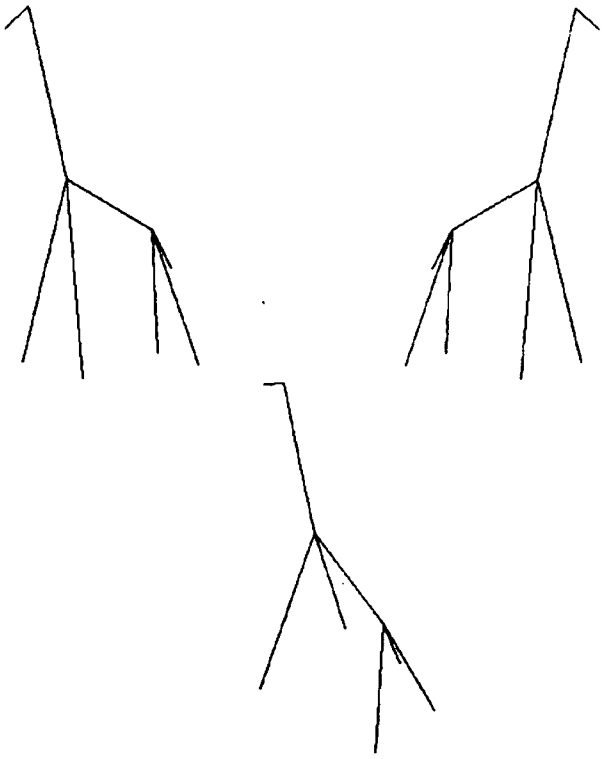

Figure 1: *Given a single 2D view (upper left), a new view (upper right) is generated under the assumption of bilateral symmetry. The two views are sufficient to verify that a novel view (second row) corresponds to the same object as the first.*

$$\mathbf{t} = B\alpha$$

which implies (since $(B^TB)^{-1}$ exists)

$$\mathbf{t} = B(B^TB)^{-1}B^T\mathbf{t} \tag{2}$$

B can then be used to check whether $\mathbf{t}$ is a view of the correct object or not, by checking whether $\|\mathbf{t} - B(B^TB)^{-1}B^T\mathbf{t}\| = 0$ or not (a further test for rigidity may also be applied, if desired, to the three available views). Figure 1 shows the results of using this technique to recognize simple pipe-cleaner animals.

2.2 Structure from One 2D Model View

Suppose, as before, that we have a single 2D view of an object. Assume further that we hypothesize (correctly) that the object is twice bilaterally symmetric (we assume in the present notation that x, y are the image coordinates and z is ortogonal to the image plane) and that symmetric quadropoles can be identified, that is sets of four points (they are the "elementary" features in this situation, since any point, which is not on both symmetry planes, corresponds to 3 other points). We define an object to be twice bilaterally symmetric if the following transformations of any 2D view of a feature quadropole yield *legal views* of the quadropole, that is orthographic projections of rigid rotations of the object:

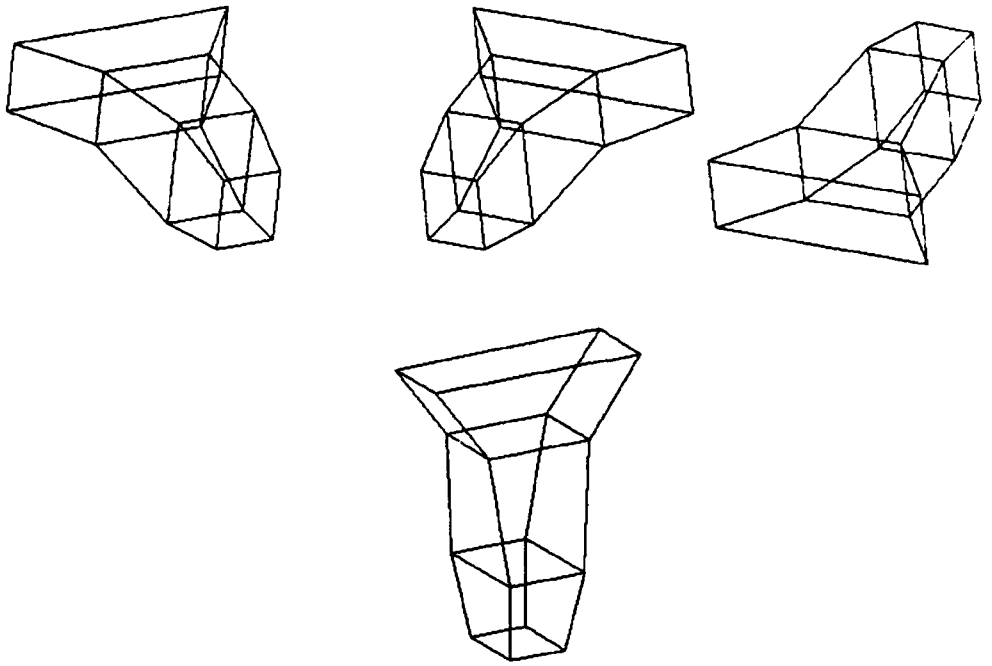

Figure 2: *A single 2D view (upper left) of a twice bilaterally symmetric object can generate additional views (upper center and right) using the symmetric properties of the object. Those three views are sufficient to compute 3D structure, as indicated in the second row where we project the 3D structure computed from the 3 views above.*

$$D_{21}\mathbf{x}_{quadr} = \mathbf{x}^1_{quadr} \tag{3}$$

$$D_{22}\mathbf{x}_{quadr} = \mathbf{x}^2_{quadr} \tag{4}$$

with

$$\mathbf{x}_{quadr} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}, \quad \mathbf{x}^1_{quadr} = \begin{pmatrix} -x_2 \\ -x_1 \\ -x_4 \\ -x_3 \\ y_2 \\ y_1 \\ y_4 \\ y_3 \end{pmatrix} \quad and \quad \mathbf{x}^2_{quadr} = \begin{pmatrix} x_4 \\ x_3 \\ x_2 \\ x_1 \\ -y_4 \\ -y_3 \\ -y_2 \\ -y_1 \end{pmatrix}.$$

These 3 views are independent apart from special views, such as accidental views (see previous section). Thus the above definition of symmetry provides a way to generate two additional views from the given one view, unless $\mathbf{x}_{quadr}$ is a view which is left invariant by at least one of the symmetry transformations D_i. This is the case, for instance, for exactly frontal views. The same argument can be repeated for all symmetric quadrupoles.

Thus, this transformations yields in the generic case to 3 independent views of the object (the symmetry yields a total of 16 views, representing 16 different orientations of the object,

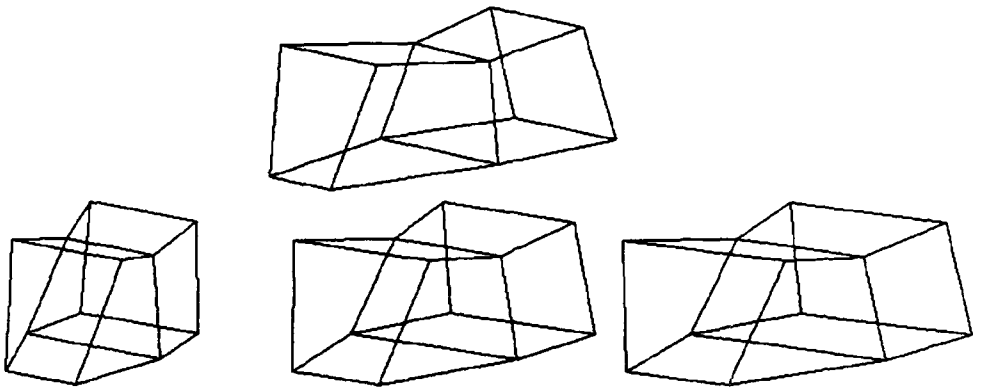

Figure 3: *A single 2D view (upper row) of a bilaterally symmetric object can be generated by different bilaterally symmetric 3D objects. The three objects projected in the second row all generate the 2D view of the first row after a rotation of* 20° *around the vertical axis.*

which span the 6 dimensional viewing space of the object). One can verify that standard structure-from-motion techniques (Huang and Lee, 1989; see also Ullman, 1979) can be applied to conclude that structure is uniquely determined up to a reflection about the image plane [3]. The following holds:

Theorem 2.2 *Given a single 2D orthographic view of a twice bilaterally symmetric object (with at least* 2 *symmetric, nondegenerate quadropole features containing a total of at least four non-coplanar points) the corresponding structure is uniquely determined up to a reflection about the image plane.*

In addition, the following results can be easily derived:

1. 3D structure can be obtained from two 2D view of a bilaterally symmetric object.
2. Structure cannot be uniquely obtained from a single 2D view of a bilaterally symmetric object. So a single 2D view of an bilaterally symmetric object can be generated by different bilaterally symmetric objects (see for example figure 3).

2.3 Correspondence and Bilateral Symmetry

Let us suppose that the correspondence of 4 non coplanar points (or more) between two views (the model view and the noval view) is given (as in A.6) and the object belongs to the class of bilaterally symmetric objects. Then the argument of Appendix 6 can be applied to each of the two views generated by the model view and the assumption of bilateral symmetry (see equation 1). For each point in the first view the corresponding point (x, y) in the second view satisfies then the two equations:

[3] The W matrix defined by Weinshall (1992) is full rank in this case. It is rank deficient for simple bilateral symmetry.

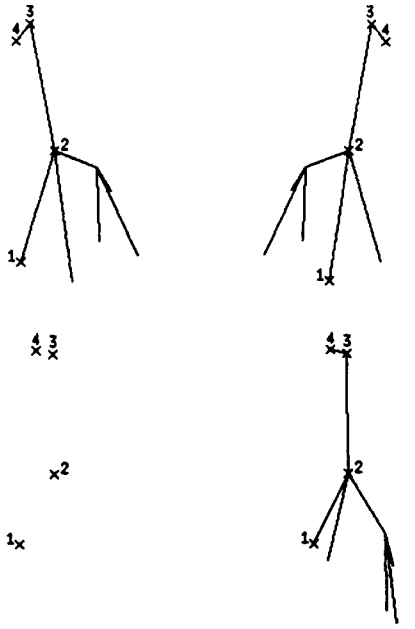

Figure 4: *Given a single 2D view (upper left), a second view (upper right) is generated by exploiting under the assumption of bilateral symmetry. Four corresponding points (lower left) are sufficient to obtain full correspondence between the model view (top left) and the novel view (lower right), of the same 3D object undergoing a uniform affine transformation.*

$$y = mx + A$$
$$y = m'x + A'$$

and is therefore uniquely determined (apart special cases) as

$$y = \frac{m'A - mA'}{m' - m} \quad , \quad x = \frac{A' - A}{m - m'} \quad .$$

Thus *the correspondence of 4 non coplanar points between two 2D views of a bilaterally symmetric object (undergoing a uniform affine transformation) uniquely determines correspondence of all other points.*

Figure 4 shows an example of obtaining full correspondence between a model view and a novel view given just four matched points and bilateral symmetry.

3 PART II: Learning Transformations

The key idea that motivated the work described in this paper is to use appropriate transformations to generate new views from a single 2D view. Such transformations may be known *a priori* as a property of the object. This is the case discussed in Part I of this paper where the symmetry properties of the objects provide the transformations. They can also be synthesized in various ways. Poggio (1991) describes a few simple techniques such as the use of 3D models whose parameters are estimated from the single 2D view. The 3D model can

be then transformed (for instance rotated) and new views thereby produced. This technique has been already used for image compression (see Aizawa, Harashima and Saito, 1989).

We are interested in a different approach. The general idea is to use an approximation technique, such as HyperBf networks, to learn an appropriate specific transformation from a set of examples of objects of the same class. For instance, we may learn a specific transformation that changes expression (from serious to smiling, say) of a face, using a set of examples consisting of pairs of 2D views of faces (each pair consists of two views of the same face, once serious and once smiling). In this section, we consider a more restricted set of transformations, *uniform affine transformations* of *2D views* of objects (see Appendix for definitions), such as rotations, in order to begin to characterize their learnability. In the case of faces one such transformation would be a specific rotation, for instance, from $+30°$ to $0°$. It is worth emphasizing that the transformations we consider here are very specific (from a ceratin specific pose to another). The situation is quite different from Part I, where we were not interested in learning transformations and we were not restricted to specific transformations.

We first introduce a very specific definition of object classes that we call *linear object classes*, for which it is easy to show existence and learnability of exact transformations. We do not believe that this is the best or most powerful definition of object classes. Its main merit is that it is simple and easy to analyse. We believe that other definitions should also be studied and that their computational and psychophysical relevance should be characterized.

3.1 Linear Object Classes

Consider a 3D view $\mathbf{X}_0$ of object 0. Assume that $\mathbf{X}_0 \in \Re^{3n}$ is the linear combination of frontal views of q 3D views of *other* objects of the same dimensionality, that is

$$\mathbf{X}_0 = \sum_{i=1}^{q} \alpha_i \mathbf{X}_i \tag{5}$$

$\mathbf{X}_0$ is then the weighted average of q points in a $3n$ dimensional space. Consider now the operator L^r associated with a desired uniform transformation (see Appendix) such as for instance a specific rotation in 3D. Let us define $\mathbf{X}_i^r = L^r \mathbf{X}_i$ the rotated 3D view of object i. Because of linearity of the group of uniform linear transformations $\mathcal{L}$, it follows that

$$\mathbf{X}_0^r = \sum_{i=1}^{q} \alpha_i \mathbf{X}_i^r.$$

Thus, *if a 3D view of an object can be represented as the weighted sum of views of other objects, its rotated view is a linear combination of the rotated views of the other objects with the same weights.* The same statement also holds for the corresponding 2D views, obtained from the 3D views under orthographic projection (see Appendix), that is

$$\mathbf{x}_0 = \sum_{i=1}^{q} \alpha_i \mathbf{x}_i \tag{6}$$

implies

$$\mathbf{x}_0^r = \sum_{i=1}^{q} \alpha_i \mathbf{x}_i^r.$$

with $\mathbf{x}_0 = P\mathbf{X}_0$, $\mathbf{x}_0^r = P\mathbf{X}_0^r$, $\mathbf{x}_i = P\mathbf{X}_i$ and $\mathbf{x}_i^r = P\mathbf{X}_i^r$.

These relations suggest that we can use "prototypical" 2D views and their known transformations to synthesize an operator that will transform a 2D view into a new 2D view when the object is a linear combination of the prototypes. Notice that the decomposition of equation 5 is always possible if $q \geq 3n$, but that in general the decomposition cannot be found uniquely for one 2D view and the given prototypes. However, if $q < 2n$, then it is possible to recover the coefficients α_i. This observation leads to:

Definition of a linear object class
A set of 3D views (of objects) $\{\mathbf{X}_i\}$ *is a* **linear object class** *if* $dim\{\mathbf{X}_i\} \leq 2n$ *with* $\mathbf{X}_i \in \Re^{3n}$.

This is equivalent to say that all objects of the same class cluster in a small linear subspace of $\Re^{2n}$ spanned by $2n$ prototypes. Edelman (1992) discusses closely related issues in the context of the complexity of recognition.

3.2 How to Learn Transformations for Linear Object Classes

First we compute the coefficients α for the optimal decomposition (in the sense of least square) of a "initial" view $\mathbf{x}_0$ of an object Φ into the "initial" views $\mathbf{x}_i$ of the q given prototypes by minimizing

$$||\mathbf{x}_0 - \sum_{i=1}^{q} \alpha_i \mathbf{x}_i||^2. \tag{7}$$

we rewrite equation 6 as

$$\mathbf{x}_0 = \Xi\alpha \tag{8}$$

where Ξ is the matrix formed by the q vectors $\mathbf{x}_i$ arranged column-wise and α is the column vector of the α coefficients. Minimizing equation 7 gives

$$\alpha = (\Xi)^+ \mathbf{x}_0 \tag{9}$$

The observation of the previous section implies that the operator that transforms $\mathbf{x}_0$ into $\mathbf{x}_0^r$ through $\mathbf{x}_0^r = L\mathbf{x}_0$, is given by

$$\mathbf{x}_0^r = \Xi^r\alpha = \Xi^r\Xi^+\mathbf{x}_0 \tag{10}$$

as

$$L = \Xi^r\Xi^+, \tag{11}$$

and thus can be learned from the 2D example pairs $(\mathbf{x}_i, \mathbf{x}_i^r)$. In this case, a one-layer, linear network (compare Hurlbert and Poggio, 1988) can be used to learn the transformation L. L

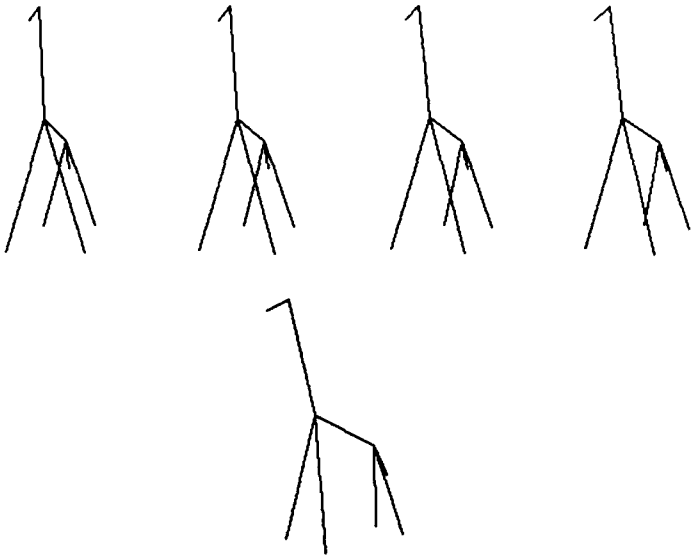

Figure 5: *Four 2D views (top) of a 3D object rotated three times around a fixed axis, each time by 5°. From the resulting 3 pairs of 2Dviews, the transformation "rotation by 5° degrees" can be learned in terms of the linear operator L. The lower row shows the effect of applying the transformation, iterated 10 times, to the upper right 2Dview.*

can then transform a view of novel object of the same class. If the q examples are linearly independent $\Xi^+ = (\Xi^T\Xi)^{-1}\Xi^T$ and the minimization of equation 7 provides $\mathbf{x}_0 = \sum_{i=1}^{q} \alpha_i \mathbf{x}_i$ [4].

3.3 Examples of Linear Object Classes and the Role of Symmetry

A "small" number of prototypes
Each set of $2n$ linear independent objects defines a linear object class, which contains all their linear combinations.

The space of a single object
As recently discovered by Basri and Ullman (1989), the space spanned by all rotations of one object has dimension 6. The dimension is reduced to 3 if the rotations are limited to the rotation around one fixed axis. A few examples (6 or 3) therefore span the complete view space in which any novel view of the same object – obtained through a 3D rotation (or any uniform affine transformation in 3D), followed by orthographic projection – lies. It is possible to transform this transformed view again, and thus to compute all the 2D views generated by a stepwise 3D rotation (see figure 5).

[4] Under more general assumptions, however, such as perspective projection and use of other non-geometric features instead or in addition to the x, y coordinates of labeled surface points, we expect the mapping between "frontal" views and rotated views to be nonlinear. Techniques such as Hyperbf should then be used (Poggio and Girosi, 1990).

Objects with a symmetry structure
3D objects with a common or partly common interior structure (e.g. symmetry, fixed angles (such as right angles) or fixed ratios between some feature points) may form a linear object class. The following result holds: *A class of objects, each one represented by a special 3D view* $\{\mathbf{X_S}\}$, $\mathbf{X}_S \in \Re^{3n}$, *is a specific linear object class if the structure can be represented by a matrix* $\mathbf{S(3n, 3n)}$ *with* $\mathbf{rank(S)} \leq \mathbf{2n}$ *and* $\mathbf{X}_S = \mathbf{SX}_S$.

Symmetric objects form a natural linear object class of this type. In the case of bilateral symmetry in the y, z plane $\{\mathbf{X_S}\}$ is taken to be a "frontal" 3D view of the object and S can be written as:

$$\mathbf{S} = \begin{pmatrix} S_{pl} & 0 \\ 0 & S_{bi} \end{pmatrix}$$

where $\mathbf{S}_{pl}(3p, 3p)$ defines the structure of p points in the symmetry plane and $\mathbf{S}_{bi}(2\times 3b, 2\times 3b)$ of b pairs of symmetric points. Both $\mathbf{S}_{pl}$ and $\mathbf{S}_{bi}$ can be written in a diagonal form

$$\mathbf{S}_{pl} = \begin{pmatrix} s_{pl} & 0 & . & 0 \\ 0 & s_{pl} & . & 0 \\ . & . & . & . \\ 0 & 0 & . & s_{pl} \end{pmatrix} \qquad \mathbf{S}_{bi} = \begin{pmatrix} s_{bi} & 0 & . & 0 \\ 0 & s_{bi} & . & 0 \\ . & . & . & . \\ 0 & 0 & . & s_{bi} \end{pmatrix}$$

where $\mathbf{s}_{pl}$ and $\mathbf{s}_{bi}$ are a square matrix

$$\mathbf{s}_{pl} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \mathbf{s}_{bi} = \begin{pmatrix} I & 0 \\ M_1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

where

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad M_1 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The final dimension d of such a class $\{\mathbf{X_S}\}$ is determined by the number p of feature points in the symmetry plane and the number b of symmetric feature pairs. Feature points in one symmetry plane reduce the dimension from $3p$ to the upper limit of $2p$. Points and their symmetric counterparts reduce the dimension from $2 \times (3b)$ to $3b$.

3.3.1 Learning the Transformation Component by Component

In the previous section we considered learning the appropriate transformation from full views. In this case the examples (prototypes) must have the same dimensionality as a full view. Our arguments above show that dimensionality determines the number of example pairs needed for a correct transformation. This section suggests that components of an object –

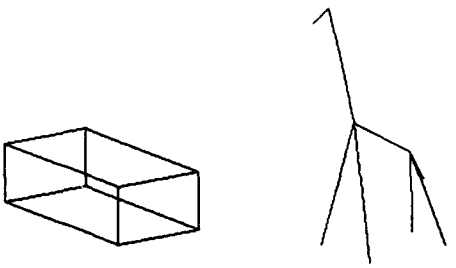

Figure 6: *Two examples of symmetric objects. The 3 symmetry planes of a cuboid reduce the effective dimensions of the object space of all cuboids from 24 to 3. The bilateral symmetry of objects consisting of 9 feature points with 5 points on the symmetry plane (see right figure) reduces the dimensions from 27 to 16.*

i.e. a subset of the full set of features – that are element of the same object class may be used to learn a single transformation with a reduced number of examples, because of the smaller dimensionality of each component. The basic components in which a view can be decomposed are given by the irreducible submatrices S_i of the structure matrix S so that $S = S_1 \oplus .... \oplus S_k$.

Consider again the linear class of *bilaterally symmetric objects.* The "diagonal" structure of S with only two submatrices is preserved after a linear transformation of the feature points in $\Re^3$:

$$L^{3n} X_S = L^{3n} S X_S$$
$$X_S^r = S^r X_S$$

This shows that the problem of transforming the 2D view $\mathbf{x}_S$ of the 3D objects $\mathbf{X}_S$ into the transformed 2D views $\mathbf{x}_S^r$, can be treated separately for each component of $\mathbf{x}_S$. For simplicity, we deal in the following only with symmetric pairs of feature points (points on the symmetry plane are degenerate pairs). The components are determined by the submatrix $\mathbf{s}_{bi}$ on the diagonal of S and are the 2D coordinates of a pair of bilaterally symmetric points $\mathbf{x}_{bi}$. The constraint $\mathbf{X}_S = \mathbf{S}\mathbf{X}_S$ leads to :

$$X_{bi} = s_{bi} X_{bi}$$

This equation is equivalent to equation 5. Therefore the linearly independent column vectors of $\mathbf{s}_{bi}$ span the 3 dimensional space of a pair of symmetric points. It follows that *3 examples are sufficient to learn a transformation of a pair of bilateral symmetric points (using a linear network).*

This observation shows the dramatic decrease in the number of examples necessary for learning the specific transformation if a bilateral symmetric object is transformed by components. In this case a single basic component consists of a symmetric pair of points. A total of 3 examples for each pair of points, are sufficient to learn a specific transformation (such as a rotation from α to β around a prespecified axis) of any bilaterally symmetric object. As shown earlier the lower limit for the number of examples is $1/2 \times 3n$ for a symmetric objects consisting of n points if the objects is transformed as a whole.

4 Concluding Remarks

- *Classifying a novel view in terms of an object class*

  We have left open the question of how to classify the object from a novel 2D view. This is the first step for then inferring certain symmetry properties or for applying learned transformations. Notice that hypotheses about symmetries can always be attempted and tried out.

- *Identifying a symmetry pair (or a n-ple)*

  The techniques of Part I require identification in the novel view of symmetry pairs (or quadropoles). Additional information may be available (e.g. once the two eyes are identified as eyes, it is known that they represent a symmetric pair). In other cases (e.g. line drawings of geometric objects) algorithms capable of identifying feature points likely to be symmetric should be feasible. Though we have not worked on this problem yet. It is intriguing to speculate about relations to known human abilities of detecting symmetries and with human tendencies of hypothesizing symmetry in visual perception.

- *Exact frontal model views should be avoided*

  Our results about bilateral symmetry imply that one should avoid to use in the data base a model view which is a fixed point of the symmetry transformations (since the transformation of it generates an identical new view). In the case of faces, this implies that the model view in the data base should not be an exactly frontal view.

- *A symmetry of higher order than bilateral allows recovery of structure from one 2D view*

  Our results imply that even when other cues that provide structure from 1 view (such as shading, perspective, texture etc.) are absent, an object symmetry of sufficiently high order may provide structure from a single view. An interesting conjecture is that human perception may be biased to impose a symmetry assumption (in the absence of other evidence to the contrary), in order to compute structure.

- *A new algorithm for computing structure from single views of polyedric objects*

  Marrill (1991) proposed an iterative algorithm that is capable of recovering structure from single views of some simple geometric solids. Sinha (1992) has improved considerably the algorithm and shown that it works well on a wide range of line drawings. Our result on structure-from-1-view may explain some of these results in terms of the underlying algebraic structure induced by symmetry properties (or other properties, for instance constraints on angles). It also yields a new non-iterative algorithm for the recovery of structure since it provides (once symmetric n-ple are identified) a simple algorithm generating a total of 3 linearly independent views to which any of the classical S-f-M algorithms can be applied, including the recent linear ones (Huang and Lee, 1989). It remains an open question to characterize the connection between the minimization principle of Marrill-Sinha and our internal structure constraints.

- *A practical algorithm for face recognition, based on features*

  Assume to have one almost-frontal image per person in the data base. The matrix B is synthesized for each person by identifying a set of symmetric pairs (eyes, etc.) and performing the operations described earlier on the model view. When a novel view is presented:

  1. Assume or infer that the image represent a face
  2. Identify pairs of symmetric points, such as the eyes
  3. Apply to the vector associated to the novel view the operator $B(B^TB)^{-1}B^T$ to verify recognition.

- *An even more practical algorithm for face recognition, based on "grey"-levels*

  Assume to have an almost-frontal quasi-grey-level image per person in the data base (Brunelli and Poggio, 1991). Assume that symmetric pairs are identified in the data base image. Assume further that four points can be found (such as one eyes, the corners of the mouth, and the top of the nose) and matched between the novel view and the model view. Then all other points (assuming that faces are sufficiently symmetric!) can be matched (disregarding self occlusions) and a distance (or correlation) measure can be computed. This technique assumes quasi-constant illumination and is not invariant to expression. It is invariant to scaling and pose (modulus self-occlusions). We are presently working towards testing and extending this basic technique. It may lead to practical applications in model acquisitions and 3D object recognition, since it makes possible to combine features and grey levels in an elegant and efficient way.

- *From views to grey-level images*

  The obvious way to go from views (see Appendix for definition) to grey-level images is through texture mapping (Poggio and Brunelli, 1990).

- *Other definitions and uses of prototypes*

  S. Ullman has suggested that it may be wiser to define – instead of the several prototypes of equation 5 – one single prototype and a small set of "perturbation" vectors. This is formally completely equivalent to the formulation of section 3.1, but it may better capture the psychophysics of object recognition.

  We should also mention, though this is somewhat outside the scope of this paper, that it is possible to use prototypes – say of a face – to compute parameters, such as illumination and pose, that may help to "normalize" a later recognition step. Poggio and Edelman (1990) used a HyperBF network to learn to associate to a 2D view of a specific object the correct 3D pose parameters. It seems that a reasonable performance may be achieved for similar tasks by using appropriate prototype(s) of the specific object (R. Basri also suggested a similar idea).

- *Nonlinear object classes and nonlinear transformations*

  The basic idea of Part II – to learn appropriate transformations from instances of the same object class – can be applied to object classes other than the linear classes we

have defined and characterized. In addition, transformations to be learned may be nonlinear or non-uniform (we have only considered linear, uniform transformations on 2D views): an example is the transformation that changes expression of a face from serious to smiling or the transformation that "ages" a face. Nonlinear object classes and nonuniform, nonlinear transformation require learning techniques more powerful than the linear ones we have considered in Part II. Approximation networks such as Hyperbf (Poggio and Girosi, 1990) may be needed.

- *An alternative to elastic templates*

  Elastic templates have been used for at least twenty years to perform recognition when only one (or very few) templates are available. Elastic templates are equivalent to using complex metrics (i.e. cost functionals) that take into account prior knowledge about allowed deformations and penalize them accordingly. Though there are techniques, such as Hyperbf (Poggio and Girosi, 1990), that can learn – to some extent – the appropriate metric (through the matrix $\mathbf{W}$) from examples, in general the art of generating good elastic templates is "black magic". In addition, elastic templates are usually very expensive computationally at run-time (because of the usually non-convex minimization problem). A more classical and formally more satisfying approach is to have a fixed metric (or almost fixed: certain invariances such as translation for which the cost is zero, if valid for the specific problem, should be embedded in the cost functional or in the choice of the input features from the very beginning) and to provide a sufficient number of examples of allowed and not allowed deformations. One could then use classification or approximation techniques such as Hbf. In some cases, however, only very few examples of deformations (or none) are readily available. The idea is then to generate artificial examples of deformations for the specific object of interest by learning the allowed deformations from a set of examples of objects of the same class, using standard approximation techniques.

A The 1.5 view theorem and other useful background math

A.1 Summary of the Appendix

This appendix [5] introduces definitions and results that characterize the algebraic structure of the views of one 3D object under orthographic projection. Consider the linear vector space $\Re^{3N}$ of 3D views of all objects, with a 3D view being the vector of the x, y and z coordinates of each of N feature points. Consider the suspace $V^{3N}_{ob_i}$ generated by one view of a specific object and by the action on it of the group of *uniform* linear transformations $\mathcal{L}$ (i.e. the same linear transformation is applied to each feature point). $\mathcal{L}$ is an algebra of order 9, and therefore a linear vector space isomorphic to $\mathcal{M}_3$ (that is the space of the 3×3 matrices with real elements). Thus, $V^{3N}_{ob_i}$ is a linear vector space isomorphic to $\Re^9$. The projection operator (orthographic projection) that deletes the z components from the 3D views, maps $V^{3N}_{ob_i}$ into a linear vector subspace $V^{2N}_{ob_i}$, isomorphic to $\Re^6$. $V^{2N}_{ob_i}$ consist of vector with x and y components and can be written as the direct sum $V^{2N}_{ob_i} = V^N_x \oplus V^N_y$, where V^N_x and V^N_y are non-intersecting linear subspaces, each isomorphic to $\Re^3$. In addition, Poggio (1990) has proved (Basri obtained this result independently, see Ullman and Basri, 1991) that $V^N_x = V^N_y$, which implies that 1.5 snapshots are sufficient for "learning" an object (generically) and performing recognition of a novel view. If 3D translations are included, a linear subspace, isomorphic to $\Re^2$ must be added to the linear space spanned by the 2D views of one object. The 1.5 *views theorem* implies that the x and the y vectors obtained from the 2 frames are linearly dependent. This in turn implies that 4 matched points across two views are sufficient (generically) to determine 1-D epipolar lines for matching all other points. This is an useful result (first obtained in a different context by Huang and Lee, 1989, see also Basri, 1991 and Shashua, 1991) in correspondence problems involving 2 frames and affine, uniform transformations in $3D$.

A.2 Introduction

Basri and Ullman (1989) have recently discovered the striking fact that under orthographic projection a view of a 3D object is the linear combination of a small number of views of the same object. In this note, we reformulate their results in the more abstract setting of linear algebra. This framework makes the result very transparent: the constraint of uniform linear transformation (the same linear transformation for each vertex) implies immediately that the set of views of an object spans a 9-dimensional linear vector space, independently of the number of vertices; orthographic projection preserves linearity while reducing the number of dimensions to 6. Simple considerations show that the linear spaces of the x and y coordinates are nonintersecting and that each has dimension 3. Furthermore it can be proved (Poggio, 1990) that they are equivalent, implying that 1.5 snapshots are sufficient to learn the model of one object. We do not consider here the additional constraint of restricting the uniform affine transformation to be rigid, i.e. to be a rotation. Rotations generate a

[5] The content of this appendix is from Poggio, 1990 (IRST Technical Report 9005–03, 1990)

nonlinear subspace of $V_{ob_i}^{3N}$. It is easy to test for rigidity; it is more difficult to understand the (nonlinear) algebraic structure (see last section in Poggio, 1990).

A.3 Any view of a 3D object is a linear combination of a small, fixed numer of views

This section provides the main result of Basri and Ullman (in the second subsection).

A.3.1 Any 3D-view of an object is a linear combination of 9 views

Let us define a *3D-view* of a specific 3D object as:

$$\mathbf{X}^{obj} = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \\ x_2 \\ y_2 \\ z_2 \\ . \\ . \\ . \\ x_n \\ y_n \\ z_n \end{pmatrix}$$

with $X \in \Re^{3n}$, which is a vector space in the usual way.

We consider the set of *uniform* (my definition) linear operators on $\Re^{3n}$, defined by the $3n \times 3n$ matrices $\mathbf{L}^{3n}$, where $\mathbf{L}^{3n} = \mathbf{I}_n \otimes L$ is the tensor product of $\mathbf{I}_n$ and L:

$$\mathbf{L}^{3n} = \begin{pmatrix} L & 0 & . & 0 \\ 0 & L & . & 0 \\ . & . & . & . \\ 0 & 0 & . & L \end{pmatrix}$$

where

$$L = \begin{pmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{pmatrix}$$

is an affine transformation on $\Re^3$. Translation in $3D$ space is taken care of separately (see later).

The space of the $\mathbf{L}^{3n}$ operators is a vector space which is *isomorphic* to the vector space of the L matrices. It therefore has a basis of 9 elements independently of n.

We can express

$$\mathbf{L}^{3n} = \sum_{i=1}^{9} a_i \mathbf{L}_i^{3n}$$

where a_i can be identified with the appropriate $l_{i,j}$ and L_i^{3n} with the usual basis for L^{3n}, i.e. with the elementary matrices E, and thus

$$\mathbf{X} = \mathbf{L}^{3n}\mathbf{X}_0 = \sum_{i=1}^{9} a_i \mathbf{L}_i^{3n}\mathbf{X}_0 = \sum_{i=1}^{9} a_i \mathbf{X}_i$$

where $\mathbf{X}_i$ are 9 independent 3D views of the specific object, needed to span the 9 elements of L, 3 for each coordinate, and X_0 is a particular view chosen as the "initial" view. Thus:

Theorem A.1 *The vector space V_{0b}^{3D} generated by the action of uniform linear transformations on a 3D view of a specific object is a 9-dimensional subspace of $\Re^{3n}$, 3 dimensions for x, 3 for y and 3 for z.*

Thus any object ob_i corresponds to a low dimensional subspace $V_{ob_i}^{3D}$ of the space of all possible views of all objects $\Re^{3n}$. Of course, $V_{ob_i}^{3D} \not\equiv \Re^{3n}$, *iff* $n > 3$. In other words, to have object specificity, i.e., for this result to be *nontrivial*, it is necessary that $n > 3$ (translations are supposed to be factored out by using an extra pair). Notice that $\Re^{3n} = V_{ob_i} + V_{ob_2} + \ldots$.

A.3.2 Any 2D-view of a 3D object is a linear combination of 6 2D-views

Now consider the orthographic projection $P : \Re^{3n} \rightarrow \Re^{2n}$, defined by $P\mathbf{X} = \mathbf{x}$, that is

$$P\begin{pmatrix} x_1 \\ y_1 \\ z_1 \\ x_2 \\ y_2 \\ z_2 \\ . \\ . \\ . \\ x_n \\ y_n \\ z_n \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ x_3 \\ y_3 \\ . \\ . \\ . \\ x_n \\ y_n \end{pmatrix}$$

with P being a linear operator with the matrix representation

$$P = \begin{pmatrix} 1 & 0 & . & . & . & . & . & . & 0 \\ 0 & 1 & 0 & . & . & . & . & . & 0 \\ 0 & 0 & 0 & 1 & 0 & . & . & . & 0 \\ . & . & . & . & . & . & . & . & . \\ . & . & . & 0 & 0 & . & 1 & 0 & 0 \\ 0 & 0 & . & . & . & . & 0 & 1 & 0 \end{pmatrix}$$

We define $\mathbf{x}$ as the *2D-view* of a 3D object.

The result below follows immediately (6 views span the elements of L in the first 2 rows) and is the main result of Basri and Ullman (in a different formulation):

Theorem A.2 *The vector space V_{ob_i} given by $V_{ob_i} = PV^{3D}_{ob_i}$ is a six-dimensional subspace of $\Re^{2n}$ (the space of all 2D orthographic views of all 3D objects), i.e.* $\mathbf{x_{ob}} = \sum_{i=1}^{6} a_i \mathbf{x}^i_{ob}$.

The inclusion of rigid translations is equivalent to the addition of a two-dimensional linear subspace (the same for all objects), spanned by the vectors

$$\mathbf{t_x} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ . \\ . \\ . \end{pmatrix}$$

and

$$\mathbf{t_y} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ . \\ . \\ . \end{pmatrix}$$

.

A.4 The x and the y coordinates of a view are each a separate linear combination of 3 views

In the previous section we have seen that any *2D-view* of a 3D object under orthographic projection is the linear combination of 6 *2D-views*. This section reformulates another observation of Ullman and Basri: the x coordinates of a *2D-view* are a linear combination of the x coordinates of 3 *2D-views* and the y coordinates are a linear combination of the y coordinates of 3 *2D-views*, the two combinations being independent of each other.

Let us consider a similarity transformation of $\mathbf{x}$:

$$T\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ . \\ . \\ . \\ x_n \\ y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \\ z_1 \\ . \\ . \\ . \\ z_n \end{pmatrix}$$

Under this similarity transformation, $\mathbf{L}^{3n}$ becomes a $3x3$ matrix of 9 (that is $3x3$) blocks. Each block is a multiple of $I \in \Re^{n,n}$ (notice the "isomorphism" to L!).

$$T^T L T = \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{pmatrix}$$

where

$$I_{11} = \begin{pmatrix} l_{11} & 0 & 0 & . & . & . \\ 0 & l_{11} & 0 & . & . & . \\ 0 & 0 & l_{11} & . & . & . \\ . & . & . & . & . & . \end{pmatrix}$$

and so on for the other blocks.
The same argument of section A.3 makes it clear that defining

$$\xi = \begin{pmatrix} x_1 \\ . \\ . \\ . \\ x_n \end{pmatrix}$$

$$\eta = \begin{pmatrix} y_1 \\ . \\ . \\ . \\ y_n \end{pmatrix}$$

the following holds:

$$\xi = \sum_{i=1}^{3} l_{1i}\xi_i$$

$$\eta = \sum_{i=1}^{3} l_{2i}\eta_i,$$

that is,

Theorem A.3 *The subspace spanned by the vectors ξ - the x components of* $\mathbf{x}$ *- which is a $n-$dimensional subspace of V_{ob}^{2D} (which is 2n-dimensional), is spanned by three views of the x coordinates of the object undergoing uniform transformations, i.e., each ξ can be represented as the linear combination of 3 independent ξ_i. The same is true for the η: each η is an independent linear combination of 3 independent η_i. Again, $n > 3$ in order for this to be non-trivial (since $\xi \equiv \Re^n$ for $n \leq 3$), once translations are factored out.*

Remark: The basis of ξ and the basis of η depend on the specific object.

A.5 V_x and V_y have the same basis, i.e. 1.5 snapshots suffice

We know from the previous sections that $V_{ob_i}^{2N} = V_x^N \oplus V_y^N$, where $dim V_x = dim V_y = 3$. A stronger property holds

Theorem A.4 (The 1.5 view theorem) $V_x = V_y$

Proof: Assume that V_x and V_y are not identical (I consider the projections of the x and y components expressed originally in the same base in V): then there is a vector $\mathbf{y}$ which is in V_y and not in V_x (or viceversa). Then we can take the 3D view that originated $\mathbf{y}$ (through orthogonal projection) and apply to it a legal transformation consisting of a rigid rotation of 90 degrees in the image plane (such a transformation is in L and therefore is legal). The x view of that 3D vector is the $\mathbf{y}$, contradicting the assumption. It follows that $V_x = V_y$.

Remarks

1. The same argument shows that $V_x = V_y = V_z$
2. The same basis of three vectors spans V_x, V_y and V_z (separately).
3. The property that the x views and the y views of the same 3D object from the same snapshot are independent is generic, since if they were dependent, a very slightly different object, differing only in the y coordinate of one vertex would have independent views (Bruno Caprile, pers. com.).
4. In general, 1.5 snapshots are sufficient to provide a basis (with $n > 3$, once translations are factored out, in order for this to be nontrivial).
5. Any 4 vectors from V_x and V_y are linearly dependent.

A.6 A corollary of the 1.5 views theorem: given four matched points, correspondance for motion or recognition is easy

A direct consequence of the above 1.5 views theorem is that the 4 vectors (from 2 orthographic views) of the **x** and **y** components of an object undergoing an uniform affine transformation in 3D (in particular a rigid transformation in 3D) are linearly dependent, that is

$$\alpha_1 \mathbf{x}_1 + \beta_1 \mathbf{y}_1 + \alpha_2 \mathbf{x}_2 + \beta_2 \mathbf{y}_2 = 0.$$

This implies that the correspondence of at least 4 non coplanar points (including translations) in two frames determines epipolar lines for the matching of all other points (the observation is due to Ronen Basri, 1991; see also Amnon Sha'shua, 1991; a similar result – but not this proof– was first obtained by Lee and Huang, 1988). This means that for each point (x_1, y_1) in frame 1, the corresponding point in frame 2 satisfies the equation

$$y = mx + A$$

with $m = -\alpha_2^*$ and $A = -(\alpha_1^* x_1 + \beta_1^* y_1)$ and $\alpha_1^* = \alpha_1/\beta_2$ and so on. Translations are taken care of by matching one point (the origin of the coordinate systems) in the two frames. Three additional "generic" points are needed to solve for α_1^*, α_2^* and β_1^*.

Therefore in problems of matching between 2 frames – in motion or recognition – four non coplanar points are sufficient to determine epipolar lines along which the matching of the other points can be more easily found.

References

[1] K. Aizawa, H. Harashima, and T. Saito. Model-based analysis synthesis image coding (mbasic) system for a person's face. *Signal Processing: Image Communication*, 1:139–152, 1989.

[2] R. Basri. On the uniqueness of correspondence under orthographic and perspective projections. A.I. Memo No. 1333, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1991.

[3] R. Basri and S. Ullman. Recognition by linear combinations of models. Technical Report CS89-11, Weizman Institute of Science, 1989.

[4] R. Brunelli and T. Poggio. Face recognition: features vs. templates. Technical Report 9110-04, I.R.S.T., Povo (IT), 1991.

[5] S. Edelman. On learning to recognize 3d objects from examples. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992.

[6] S. Edelman and T. Poggio. Bringing the grandmother back into the picture: a memory-based view of object recognition. A.I. Mem· 1181, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1990.

[7] T.S. Huang and C.H. Lee. Motion and structure from orthographic projections. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(5):536–540, 1989.

[8] A. Hurlbert and T. Poggio. Synthetizing a color algorithm from examples. *Science*, 239:482–485, 1988.

[9] T. Marill. Emulating the human interpretation of line-drawings as three-dimensional objects. *International Journal of Comuter Vision*, 6:147–161, 1991.

[10] Y. Moses and S. Ullman. Limitations of non model-based recognition schemes. A.I.Memo 1301, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1991.

[11] T. Poggio. 3D object recognition: on a result by Basri and Ullman. Technical report # 9005–03, IRST, Povo, Italy, 1990.

[12] T. Poggio. 3D object recognition and prototypes: one 2D view may be sufficient. Technical Report 9107–02, I.R.S.T., Povo, Italy, July 1991.

[13] T. Poggio and S. Edelman. A network that learns to recognize 3D objects. *Nature*, 343:263–266, 1990.

[14] T. Poggio and F. Girosi. Networks for approximation and learning. *Proceedings of the IEEE*, 78(9), September 1990.

[15] Amnon Shashua. Correspondence and affine shape from two orthographic views: Motion and recognition. A.I. Memo 1327, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1991.

[16] Pawan Sinha. The apparent non-rigidity of rigid objects in motion. Technical report, Massachusetts Institute of Technology, 1992. (preprint).

[17] S. Ullman and R. Basri. Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:992–1006, 1991.

[18] Shimon Ullman. *The Interpretation of Visual Motion.* MIT Press, Cambridge and London, 1979.

[19] D. Weinshall. Model based invariants, and their use for representation, constant-time indexing, and linear structure from motion. IBM Computer Science Research Report. RC 17505 (No. 77262), I.B.M., 1992.